

The simulation of bovine serum albumin vibration spectrum in THz domain

M. MERNEA, O. CALBOREAN, M. P. DINCA^a, A. LECA^b, D. APOSTOL^b, T. DASCALU^b, D. MIHAILESCU*

University of Bucharest, Faculty of Biology

^aUniversity of Bucharest, Faculty of Physics

^bNational Institute for Laser, Plasma and Radiation Physics

Biomolecules have distinct signatures in THz spectral region. Here we reconstructed the THz theoretical spectrum of hydrated bovine serum albumin (BSA), a protein with unknown 3D structure. The THz spectrum of dry BSA was simulated using a 3D model of the protein. THz theoretical spectra of water were calculated using four water models. By adding the dry BSA and water theoretical spectra we obtained a good correlation with the experimental data that were recorded on a 10% w/w BSA solution.

(Received December 15, 2009; accepted January 20, 2010)

Keywords: THz spectroscopy, Normal modes analysis, Molecular dynamics, Spectra simulation; modelling

1. Introduction

Terahertz (THz) spectroscopy is a technique of special interest in protein spectroscopy because it covers the region of bimolecular collective movements, including high amplitude movements of the whole protein or its subdomains. THz absorption spectroscopy is able to probe these modes directly in unfolded, partially folded, and folded proteins [1]. It also can be used in a 'dynamical' mode to study directly such motions and in a 'kinetic' mode where these motions report on even larger scale complete folding and unfolding of biomolecules occurring on a 10 ps or longer time scale [2]. Theoretical THz vibration spectra can be simulated by normal modes analysis [3,4] (NMA) or by molecular dynamics (MD) [5].

Bovine serum albumin (BSA) is the main protein of bovine plasma. It is involved in the regulation of blood osmotic pressure and in the binding of metal ions, proteins or lipids. Even if BSA has been previously characterized by THz spectroscopy [6], the challenge about this protein is modelling its THz spectrum because the 3D structure of BSA is unknown.

We obtained spectra with a spectral resolution of 7.4 GHz in 0.2-2.8 THz domain for solvated BSA (10% solution) using a THz-TDS experimental setup.

Our aim was to reconstruct the THz spectrum of solvated BSA by adding the water spectrum to the BSA theoretical spectrum. We computed the THz theoretical spectrum of dry BSA using a 3D model of the protein that we have previously built by homology modelling [7]. BSA spectrum was simulated by normal modes analysis (NMA) and by molecular dynamics (MD). We have also simulated the THz spectrum of water using four theoretical water models in order to establish which water model produces results in agreement with the experimental data reported by other studies. Finally, we reconstructed the THz spectrum of solvated BSA by adding the water and BSA theoretical spectra and compared our theoretical results with our experimental data.

2. THz-TDS method and experimental setup

Over the THz range the optical parameters of biological materials are found via complex transfer function determined over a material sample. This transfer function is deduced from the modification suffered by THz pulses during sample crossing. First, a reference measurement of THz pulse intensity, $E_{\text{ref}}(t)$, is done without a sample between THz emitter and THz detector (Fig. 1(a)). Then, the second measurement records the THz intensity, $E_{\text{sample}}(t)$, when the sample is inserted between detector and emitter (as in Fig. 1(b)). The detected signal is proportional to the electric field of the terahertz pulse and, by changing the delay between the pumping and gating laser pulses, the terahertz electric field as a function of time can be recorded. A simple Fourier transformation gives the frequency spectrum. The complex transfer function of the sample is obtained

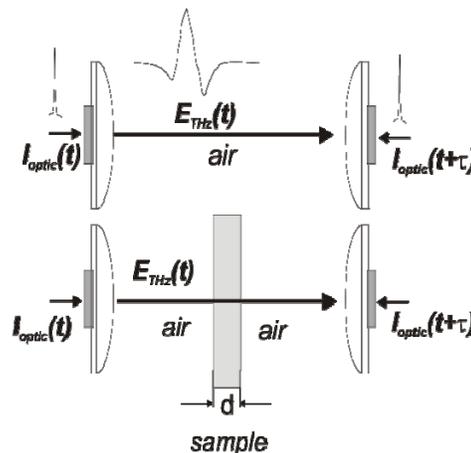


Fig. 1. The principle of THz TDS measurement.

by dividing the sample frequency spectrum to reference spectrum, $\bar{A}(\omega) = E_{sam}(\omega)/E_{ref}(\omega)$. It is essential to note that this function accurately provides both the Fourier phase and amplitude of waveforms and, consequently, the complex material properties, such as the complex refractive index and the dielectric constant, can be obtained without the complex Kramers–Kronig analysis. At the same time the temporal origins of reference and sample signal must be accurately synchronized in order to avoid an additional phase contribution in the Fourier transform. Thus, denoting by $A = \sqrt{R^2 + I^2}$ and $\varphi = \arctan(I/R)$ the magnitude and phase of the ratio of the two complex spectra $R(\omega) + jI(\omega) = E_{sam}(\omega)/E_{ref}(\omega)$, the sample index of refraction can be computed as $n(\omega) = 1 + \frac{c}{\omega d} \varphi(\omega)$

$$\text{and } k(\omega) = -\frac{c}{\omega d} \ln \left(\frac{(n(\omega)+1)^2}{4n(\omega)} A \right) \quad [8].$$

The absorption coefficient α is correlated to k by:

$$\alpha(\omega) = 2 \frac{\omega k(\omega)}{c} = -\frac{2}{d} \ln \left(\frac{(n(\omega)+1)^2}{4n(\omega)} A \right)$$

which becomes $\alpha(\omega) = -\frac{2}{d} \ln A$ if the effect of the entering interface into the sample medium is neglected. The diagram of the experimental setup for THz-TDS is presented in Fig. 2. A femtosecond (fs) pulse train was emitted from a self mode locked fiber laser (TOPTICA GmbH) SHG having the following specifications: 780nm wavelength, 80mW average power, 150 fs pulse length.

The optical pulse train having a 90 MHz repetition rate is separated in two equivalent beams by a beam-splitter. One beam travels to the THz source, a DC biased microstrip photoconductive antenna fabricated on low-temperature grown GaAs (LT-GaAs) 400 μm thick substrate and is focused by an integrated silicon lens

between the electrodes where THz pulses are produced when optical excitation induces conductivity changes in the semiconductor.

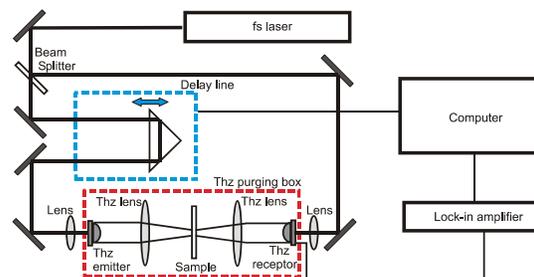


Fig. 2. TDS-THz setup for pulse transmission measurements.

The maximum bias voltage was 80 V and the standard emitted THz radiation power, when pumped by mode-locked ultra fast laser with 80 mW output power, is 10 μW . The THz pulse is then focused using polyethylene lens onto the sample, where the interaction between sample and THz radiation may alter the THz pulse shape. The beam emerging from the sample is expanded by another polyethylene lens and directed onto the THz detector with a structure similar to those of the emitter but working as an unbiased antenna, the photoconductive gap of the antenna being gated by the fs optical pulses of the second beam provided by the beam splitter. The THz pulse provides the bias voltage accelerating the charge carriers to create a current that is amplified and recorded. A delay line based on hollow retro reflector was placed in the pumping beam path to assure pulses' synchronization and allow a temporal scan of THz pulses by reconstructing it from time gate measurements. The line was driven by a computer controlled step-motor having the step length of 1.25 μm . The distance between emitter and receiver is about 40 cm. To accommodate the BSA sample a special Teflon holder has been designed.

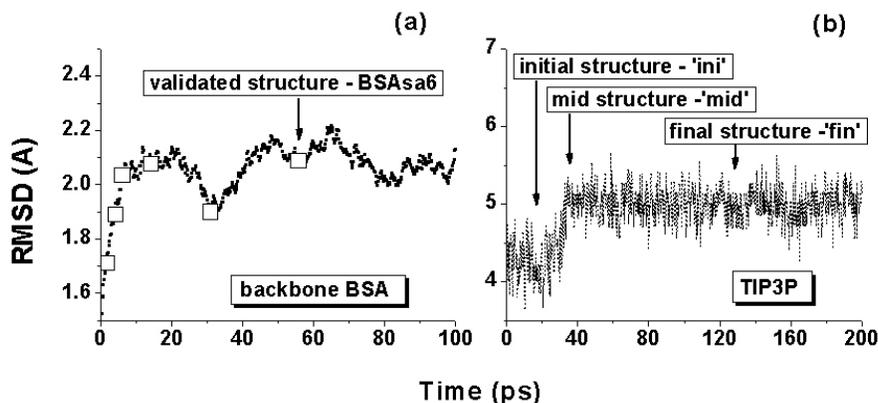


Fig. 3. (a) Structure fluctuations during the slow cool stage from the simulated annealing dynamics. BSA conformation extracted for further analysis are marked with empty squares. The validated structure is labelled. (b) Structure fluctuations of the TIP3P water box during the molecular dynamics simulation. The structures extracted for 'ab initio' calculations are labelled.

3. Theoretical methods

3.1 3D model of BSA

The 3D model of BSA was built with the homology modeling program MODELLER [9], using the known structure of human serum albumin (HSA, pdb code: 1ao6 [10]) as template (BSA - HSA amino acids sequence identity $\approx 76\%$). The amino acids sequence of BSA (UniProtKB/Swiss-Prot code P02769 [11]) was aligned by hand with the sequence of HSA that was extracted from the available coordinates file. The model of BSA was further refined by adding hydrogen atoms and disulphide bridges using the molecular simulation program CHARMM [12]. We included the structure in a rectangular water box of 24 653 water molecules using 'convpdb.pl' [13]. Energy minimization and molecular dynamics simulations of this system have been performed using NAMD [14] with CHARMM22 force-field [15] parameters. NPT dynamics with periodic boundary conditions was run with a time step of 2 fs for 1 ns. The target temperature of the Langevin thermostat was 300 K, with a damping coefficient of 1 ps^{-1} . The target pressure of 1 atmosphere was controlled by the Nosé-Hoover Langevin barostat [16] with a piston oscillation period of 200 fs and a damping time of 500 fs. Electrostatic interactions were computed using the PME [17, 18] method with a grid spacing of 1 Å. We have analyzed the temporal evolution of the system in both situations and extracted ten different BSA conformations assumed during the dynamics simulation (BSA1 – BSA10). We performed simulated annealing dynamics [19] of the hydrated initial BSA model. We extracted other six conformations assumed during the slow cool phase of the simulated annealing dynamics (BSAsa1 – BSAsa6). We also extracted the water molecules at 2.6 Å (one water layer) around each of the sixteen structures.

In a previous study we presented the method that we used to validate only a single BSA conformation. The validated structure is BSAsa6 (Fig. 3(a)). We calculated the theoretical THz spectra of dry BSAsa6 and of BSAsa6 with one water layer by NMA [20]. We also calculated the theoretical spectrum of BSA from the dynamics trajectory. Both methods are presented further in the text.

3.2 Water model test

CHARMM force field provides several water models. In order to calculate the theoretical spectrum of water we have chosen the TIP3P [15], TP3M [15, 21], WAT (from the topology for CHARMM 22 version) and TIP4P [22, 23] water models. These models have been used to build water boxes with 5 water molecules on each side. TIP3P, TP3M and WAT are three-point water models as they represent a rigid water molecule with three interaction sites. For the TIP3P, TP3M and WAT water boxes we simulated constant pressure and temperature (NPT) leap verlet dynamics for 200 ps, with periodic boundary conditions (pbc) with CHARMM [12]. Temperature and

pressure were held constant at 300 K and 1 atm using the CPT algorithm. Hydrogen bonds were held by shake and the time step was 2 fs. For the four-point water model (TIP4) simulation we have used the charge equilibration method [22, 23]. We simulated 200 ps NPT leap dynamics at 300K and 1 atm with pbc. Particle Mesh Ewald (PME) [17, 18] was used to account for the long-range electrostatics. Temperature was held constant using Nose [16] temperature bath and pressure was maintained using the CPT algorithm. The dynamics trajectories of each water box have been used for theoretical THz water absorption spectra.

For a better understanding of water spectrum at low frequencies we also calculated the intensity of water absorption in THz domain using quantum mechanical methods. We extracted three TIP3P water box structures from the dynamics trajectory after 16 ps ('ini' box), after 35 ps ('mid' box) and after 128 ps ('fin' box) (Fig. 3(b)). Then, from each structure we extracted the water molecules that could fit inside a sphere with the radius of 4.5 Å, 6 Å, 6.5 Å and 7 Å, accounting for 15, 28, 32 and 49 water molecules, resulting in 16 structures. The low frequency (THz) vibrational modes of these water clusters were calculated using the Jaguar quantum chemistry package (Schrodinger, Inc.) at RHF/6-311G** level of theory. The resulting frequencies were adjusted by using the uniform scaling factor of 0.9051 for the RHF/6-311G** level of theory recommended by the Jaguar package.

3.3 THz spectrum calculation using NMA

Prior to generating the normal modes of vibration [20], the structures were energy minimized for 500 steps performed by the adopted basis Newton-Rhpson algorithm. The minimization method was adjusted taking into account the mean square deviation of the structure during minimization and the frequencies outputted by the generation of the normal modes. THz vibrational frequencies of the structures were obtained from the normal mode eigenvalues of the mass-weighted Hessian. The intensities of each normal mode were calculated from the dipole derivatives as given by the following formula:

$$I_i = \left| \sum_{\alpha=1}^{3N} \frac{\partial \mu_{\alpha}}{\partial r_{\alpha}} X_{ij} \right|^2,$$

where the sum is over all 3N coordinates, represented as r_j and the X_{ij} are the corresponding eigenvectors derived from the diagonalization of the mass-weighted Hessian. For the comparison with the experiment, a Lorentzian function was used to describe each intensity line, as given by the formula:

$$I(\omega_i) = \frac{I_{max} \Gamma^2}{[\Gamma^2(\omega - \omega_i)^2 + \Gamma^2]} \quad [24].$$

In this case, the lines are uniformly broadened to a temperature of 300 K and Γ value (full width at half max) was set 20 cm^{-1} .

3.4 THz spectrum calculation using MD

Trajectory analysis was performed using VMD [25]. The outputted THz spectrum represents the fast Fourier transform of the real-time dipole-dipole autocorrelation function calculated for the system - $C(\omega)$. The final THz absorption spectrum is there calculated by evaluating

$$A(\omega) = \omega [1 - \exp(-\omega(kT))] C(\omega) \quad [26],$$

where k is the Boltzmann constant and T is the absolute temperature.

4. Results and discussions

4.1 THz spectrum of water

We simulated the THz spectra of water using four water models: TIP3P, TP3M, WAT and TIP4P using molecular dynamics methods (Fig. 4(a), 4(b)). TIP3P, TP3M and WAT water models are better correlated with

the experimental data reported by other studies [27-29] than the four-point water model. Among the three-point water models, the absorption spectrum of TIP3P water box reproduces better the experimental liquid water absorption peaks, as it shows peaks at 0.9, 1.2, 1.4, 2 and 2.5 THz.

It is known that water is a polar molecule and absorbs heavily in THz domain. Because recording the THz absorption spectrum of liquid water raises technical difficulties (the transmitted power decreases rapidly with increasing frequency), we compared our theoretical results with results reported by previous far infrared [27-29] studies. These studies have identified that at room temperature, liquid water absorption spectrum shows peaks around the values 0.9 THz [27], 1.35 - 1.5 THz [28, 29], 1.8 THz, 2 THz and 2.55 THz [29]. At room temperature, the liquid water spectrum also shows a barely discernable shoulder at 1.2 THz [29].

We calculated the absorption intensities of TIP3P water boxes extracted from the dynamics trajectory of the initial TIP3P water box using quantum mechanics methods. The structures investigated comprised 15, 28, 32 and 49 water molecules with different 3D arrangements. By superimposing the intensities computed for all the structures we obtain a TIP3P water spectrum in agreement with the experiment (Fig. 4(c), 4(d)).

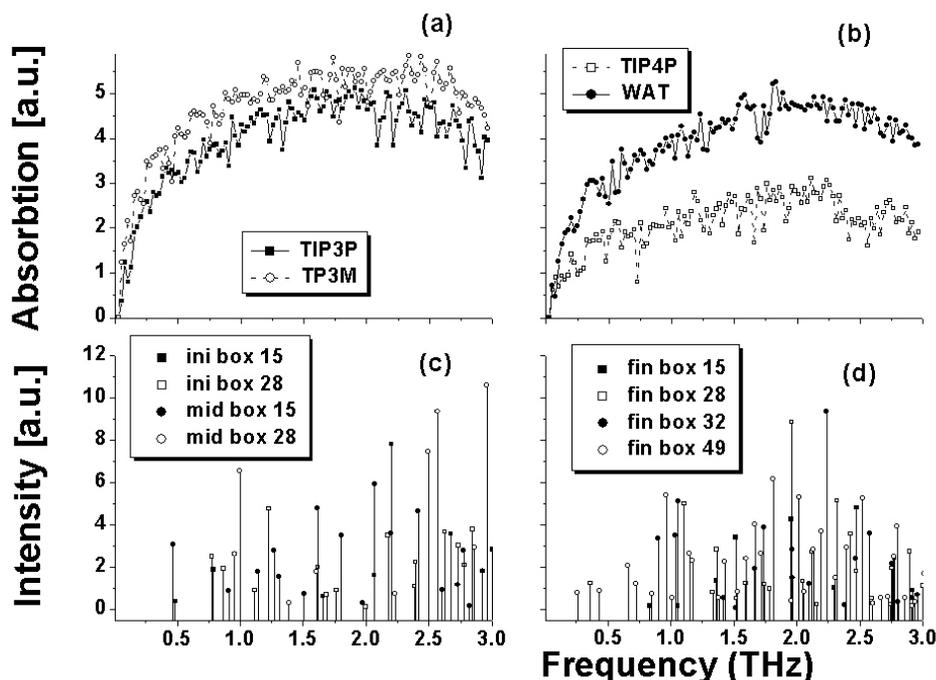


Fig. 4. (a). Theoretical THz spectra of the water boxes built using TIP3P and TP3M water models. (b). Theoretical THz spectra of the water boxes built using TIP3P and TP3M water models. (c), (d) Absorption intensity calculated for water boxes extracted from the TIP3P water dynamics trajectory.

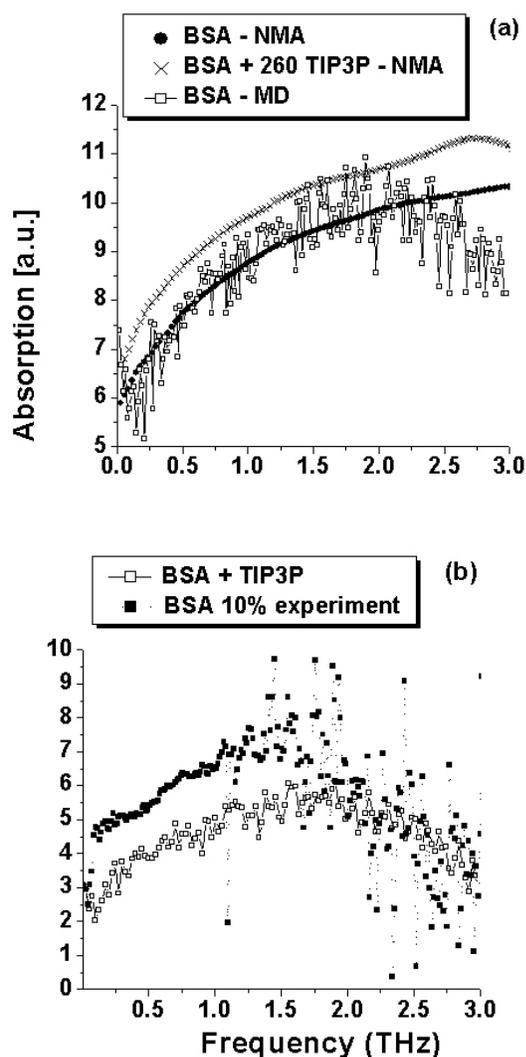


Fig. 5. (a) Representation of theoretical spectra of dry BSA derived from NMA and MD calculations. We also plotted the theoretical THz spectrum with one hydration layer (260 water molecules). (b) Representation of the experimental and theoretical spectra of solvated BSA.

4.2 THz spectrum of BSA

Using the described setup and data processing method, spectra for solvate BSA (10%) were obtained with a spectral resolution of 7.4 GHz in 0.2-2.8 THz domain.

We calculated theoretical spectra of dry BSA using NMA and MD calculations. These spectra show a good correlation in 0 to 2 THz frequency range (Fig. 5(a)). We simulated the THz of BSA with the first hydration layer comprising 260 water molecules using only NMA. The presence of the water molecules induces discrete peaks at 1.4 and 2.7 THz.

The theoretical spectrum of solvated BSA was calculated by adding the computed TIP3P water to dry BSA theoretical spectrum. We compared our theoretical data to the experimental spectrum of a BSA 10% w/w solution. We found a good agreement between theory and experiment for 0 to 2 THz (Fig. 5 (b)).

The complete structure of BSA comprises 607 residues. We have modelled ~95% of BSA accounting for 597 residues. The structure is compact, formed by α -helices connected by extended regions and reinforced by 17 disulfide bridges.

Biomolecules have distinct signatures in the THz spectral region; time-domain THz spectroscopy (THz-TDS) was successfully used to identify different molecular species, different mutations within a single species and different conformations of a given molecule [30]. Here, we used THz spectroscopy to validate our 3D model of BSA and our theoretical spectrum of solvated BSA.

5. Conclusions

Here, we report the use of THz spectroscopy for theoretical model validation and we present the method of simulating the THz spectra of solvated BSA. Using NMA and MD simulations we computed the theoretical vibration spectrum of dry BSA. Theoretical THz spectra of liquid water were computed using MD simulations and ‘ab initio’ methods. The results of the dynamics simulation and confirmed by ‘ab initio’ simulations show that the TIP3P water model is the most adequate for THz spectra calculations. By adding the dry BSA and TIP3P water theoretical spectra we calculated the theoretical spectrum of solvated BSA. Our results show a good correlation with the experimental THz data recorded on a 10% w/w BSA solution in the frequency range of 0 to 2 THz.

References

- [1] S. W. A. Markelz, J. Hillebrecht and R. Birge, *Phys. Med. Biol.* **47**, 3797 (2002).
- [2] D. F. Plusquellic, K. Siegrist, E. J. Heilweil, and O. Esenturk, *Chemphyschem* **8**, 2412 (2007).
- [3] R. Balu, H. Zhang, E. Zukowski, J. Y. Chen, A. G. Markelz, and S. K. Gregurick, *Biophys J* **94**, 3217 (2008).
- [4] D. R. Nutt, M. Meuwly, *Biophys J* **85**, 3612 (2003).
- [5] T. Globus, A. Bykhovski, T. Khromova, B. Gelmont, L. K. Tamm, L. C. Salay, in *Terahertz Physics, Devices, and Systems II*, Vol. 6772, SPIE, Boston, MA, USA, 2007, p. 67720S.
- [6] J. Xu, K. W. Plaxco, S. J. Allen, *Protein Sci* **15**, 1175 (2006).
- [7] M. Mernea, A. Leca, O. Calborean, M. P. Dinca, T. Dascalu, D. Mihailescu, in *TERA - MIR 2009 NATO Advanced Research Workshop on Terahertz and Mid Infrared Radiation: Basic Research and Applications*, Institute of Theoretical and Applied

- Physics Turunç-Marmaris, Turkey, 2009, p. 49.
- [8] B. Fischer, M. Hoffmann, H. Helm, R. Wilk, F. Rutz, T. Kleine-Ostmann, M. Koch, P. Jepsen, *Opt Express* **13**, 5205 (2005).
- [9] A. Sali, T. L. Blundell, *J Mol Biol* **234**, 779 (1993).
- [10] S. Sugio, A. Kashima, S. Mochizuki, M. Noda, K. Kobayashi, *Protein Eng* **12**, 439 (1999).
- [11] K. Hirayama, S. Akashi, M. Furuya, K. Fukuhara, *Biochem Biophys Res Commun* **173**, 639 (1990).
- [12] B. R. E. Brooks, S. D. J. Olafson, S. Swaminathan, M. Karplus, *J. Comp. Chem.* **4**, 187 (1983).
- [13] M. Feig, J. Karanicolas, C. L. Brooks, 3rd, *J Mol Graph Model* **22**, 377 (2004).
- [14] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, K. Schulten, *J Comput Chem* **26**, 1781 (2005).
- [15] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorcikiewicz-Kuczera, D. Yin, M. Karplus, *Journal of Physical Chemistry B* **102**, 3586 (1998).
- [16] S. E. Feller, Y. H. Zhang, R. W. Pastor, B. R. Brooks, *Journal of Chemical Physics* **103**, 4613 (1995).
- [17] T. Darden, D. York, and L. Pedersen, *Journal of Chemical Physics* **98**, 10089 (1993).
- [18] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *Journal of Chemical Physics* **103**, 8577 (1995).
- [19] C. A. Laughton, *Protein Engineering* **7**, 235 (1994).
- [20] M. M. Tirion, *Phys Rev Lett* **77**, 1905 (1996).
- [21] A. H. Thomas, *Journal of Computational Chemistry* **17**, 490 (1996).
- [22] P. Sandeep L. B. Charles, III, *Journal of Computational Chemistry* **25**, 1 (2004).
- [23] P. Sandeep, D. M. Alexander, Jr., L. B. Charles, III, *Journal of Computational Chemistry* **25**, 1504 (2004).
- [24] S. Jaaskelainen, C. S. Verma, R. E. Hubbard, P. Linko, L. S. Caves, *Protein Sci* **7**, 1359 (1998).
- [25] W. Humphrey, A. Dalke, K. Schulten, *J Mol Graph* **14**, 33 (1996).
- [26] M. P. Allen, D. J. Tildesley, *Computer simulation of liquids*, Clarendon, 1987 (1989 [printing]).
- [27] J. K. Vij, F. Hufnagel, *Chemical Physics Letters* **155**, 153 (1989).
- [28] J. B. Hasted, S. K. Husain, F. A. M. Frescura, J. R. Birch, *Chemical Physics Letters* **118**, 622 (1985).
- [29] K. N. Woods, H. Wiedemann, *Chemical Physics Letters* **393**, 159 (2004).
- [30] A. Markelz, S. Whitmire, J. Hillebrecht, R. Birge, *Phys Med Biol* **47**, 3797 (2002).

*Corresponding author: dan.mihailescu@bio.unibuc.ro